

- 1 1. A method of fulfilling an information need employing an index stored on a
2 computer-readable medium and comprised of preanalyzed contexts of terms
3 appearing within a plurality of documents, comprising the steps of:
4 receiving a query comprised of one or more fully specified terms and an
5 information need, wherein the information need is represented by one or
6 more at least partially unspecified terms;
7 identifying contexts in the index that contain the one or more fully
8 specified terms and zero or more at least partially unspecified terms; and
9 locating one or more matches for the information need within the
10 identified contexts.
- 11 2. The method of claim 1, further comprising the step of:
12 identifying documents in the index that contain the one or more at least
13 partially unspecified terms.
- 14 3. The method of claim 1, wherein the locating step further comprises:
15 converting the query into a finite state machine; and
16 matching the finite state machine against the identified contexts.
- 17 4. The method of claim 3, wherein the finite state machine is a finite state
18 transducer.
- 19 5. The method of claim 3, wherein the finite state machine allows for the appearance
20 of fully specified and at least partially unspecified terms in any order in a
21 potential matching context.
- 22 6. The method of claim 3, wherein the finite state machine allows for one or more
23 intervening words between the fully specified and at least partially unspecified
24 terms in a potential matching context.
- 25 7. A method of fulfilling an information need employing an index stored on a
26 computer-readable medium and comprised of preanalyzed contexts of terms
27 appearing within a plurality of documents, information indicating category
28 restrictions that the terms and contexts satisfy, and identifiers of the documents
29 and contexts containing the terms, comprising the steps of:

1 receiving a query comprised of one or more fully specified terms and an
 2 information need and at least a partial restriction on the order that the one
 3 or more fully specified terms and the information need may appear in a
 4 potential matching context, wherein the information need is represented by
 5 one or more at least partially unspecified terms reflecting a category
 6 restriction;
 7 identifying contexts in the index that contain the one or more fully
 8 specified terms and the one or more at least partially unspecified terms in
 9 the specified order;
 10 locating one or matches for the information need within the identified
 11 contexts.

- 12 8. The method of claim 7, further comprising the step of:
 13 identifying documents in the index that contain the one or more fully
 14 specified terms and the one or more at least partially unspecified terms in
 15 the specified order.
- 16 9. The method of claim 7, wherein the locating step further comprises:
 17 converting the query into a finite state machine; and
 18 matching the finite state machine against the identified contexts.
- 19 10. The method of claim 9, wherein the finite state machine is a finite state
 20 transducer.
- 21 11. The method of claim 9, wherein the finite state machine assigns a dissimilarity
 22 weight to appearances of fully specified and at least partially unspecified terms in
 23 potential matching contexts in orders other than that specified in the query.
- 24 12. The method of claim 9, wherein the finite state machine allows for one or more
 25 intervening words between the fully specified and at least partially unspecified
 26 terms in a potential matching context.

- 1 13. A method of fulfilling an information need employing an index stored on a
2 computer-readable medium and comprised of preanalyzed contexts of terms
3 appearing within a plurality of documents, comprising the steps of:
4 receiving a query comprised of one or more fully specified terms and an
5 information need, wherein the information need is represented by one or
6 more at least partially unspecified terms;
7 converting the query into a Boolean expression;
8 identifying contexts identifiers satisfying the Boolean expression; and
9 locating one or more matches for the information need within the
10 identified contexts.
- 11 14. The method of claim 13, wherein the identifying step further comprises the steps
12 of:
13 identifying document identifiers satisfying the Boolean expression.
- 14 15. The method of claims 1, 7 or 13, further comprising the steps of:
15 converting the query into a finite state machine; and
16 matching the finite state machine against the contents of the contexts
17 associated with the identified context identifiers.
- 18
- 19 16. The method of claims 1,7 or 13, wherein the contexts are stored as finite state
20 machines.
- 21 17. The method of claims 1, 7 or 13, wherein the documents are accessible over the
22 Internet.
- 23 18. The method of claims 1, 7 or 13, wherein the documents comprise World Wide
24 Web pages.
- 25 19. The method of claims 1, 7 or 13, further comprising the step of:
26 accumulating information about the one or more matches as they are
27 located.
- 28 20. The method of claim 1, further comprising the step of:
29 assigning a score to a match.

- 1 21. The method of claim 20, wherein the score reflects the number of times an
2 instance of the match is located among the plurality of documents.
- 3 22. The method of claim 1, further comprising the step of:
4 outputting one or more of the matches, or a portion thereof, thereby
5 providing a result for the query.
- 6 23. The method of claim 22, further comprising the step of:
7 outputting identifiers or locations of one or more of the documents that
8 contains a match or portion thereof that was output in the outputting step.
- 9 24. The method of claim 23, wherein a location or a document comprises a uniform
10 resource locator.
- 11 25. The method of claim 23, further comprising the step of:
12 ranking the documents that contain a match, and wherein the second
13 outputting step comprises outputting the document identifiers or locations
14 of the documents that contain a match in an order based on the ranking.
- 15 26. The method of claim 25, wherein the ranking step comprises ranking a document
16 based on the number of times a match is located within the document.
- 17 27. The method of claim 1, wherein the category restriction comprises a
18 morphological feature.
- 19 28. The method of claim 1, wherein the category restriction comprises a syntactic
20 feature.
- 21 29. The method of claim 1, wherein the category restriction comprises a computer
22 program.
- 23 30. The method of claim 1, further comprising the step of:
24 storing a match or a portion thereof.
- 25 31. The method of claim 30, further comprising the step of:
26 storing a score for the match or portion thereof.
- 27 32. The method of claim 1, further comprising the step of:
28 storing a plurality of matches or portions thereof.
- 29 33. The method of claim 1, further comprising the step of:

- 1 storing a score for a plurality of matches or portions thereof.
- 2 34. The method of claim 1, wherein the index comprises locations of terms within
- 3 documents.
- 4 35. The method of claim 34, wherein the locating step further comprises:
- 5 determining the location of a term in the query within a document using
- 6 the index; and
- 7 locating a match for the query based on the location of the term within the
- 8 document.
- 9 36. The method of claim 1, further comprising the step of:
- 10 ranking a plurality of the located matches or portions thereof.
- 11 37. The method of claim 36, wherein the ranking step comprises:
- 12 ranking a located match or a portion thereof based on the content of a
- 13 plurality of documents identified in the identifying step.
- 14 38. The method of claim 36, wherein the ranking step comprises:
- 15 ranking a located match or a portion thereof based on the content of a
- 16 majority of documents identified in the identifying step.
- 17 39. The method of claim 36, wherein the ranking is based on one or more features
- 18 selected from the list consisting of: the location of a match within a document, a
- 19 weight assigned to a document that contains a match, the age of a document that
- 20 contains a match, the source of a document that contains a match, and a format
- 21 feature of a match within a document.
- 22 40. The method of claim 36, wherein the ranking step comprises:
- 23 ranking a located match or a portion thereof based on the number of times
- 24 an instance of the match is located within a plurality of documents
- 25 identified in the identifying step.
- 26 41. The method of claim 36, wherein the ranking step comprises:
- 27 ranking a located match or a portion thereof based on the number of times
- 28 an instance of the match is located within a majority of documents
- 29 identified in the identifying step.

- 1 42. The method of claims 36, 38 or 39, further comprising the step of:
2 outputting one or more of the located matches, or one or more portions
3 thereof, in an order based on the ranking, thereby providing a result for the
4 query.
- 5 43. The method of claim 42, further comprising the step of:
6 outputting an indication of the ranking of a located match or portion
7 thereof.
- 8 44. A method of fulfilling an information need based on documents stored on a
9 computer-readable medium comprising the steps of:
10 storing an index identifying documents containing terms;
11 storing contexts for terms, wherein a context occurs in a document;
12 receiving a query containing an unspecified portion; and
13 identifying one or more matches for the query within the contexts.
- 14 45. A method of fulfilling an information need based on documents stored on a
15 computer-readable medium comprising the steps of:
16 storing an index identifying documents containing terms;
17 storing contexts, wherein a context occurs in a document;
18 storing information to retrieve a list of contexts for terms;
19 receiving a query containing an at least partially unspecified portion; and
20 identifying one or more matches for the query within the contexts.
- 21 46. The method of claim 45, wherein contexts are preanalyzed.
- 22
23 47. A method of fulfilling an information need based on documents and an index
24 stored on a computer-readable medium comprising the steps of:
25 storing contexts for terms, wherein the context occurs in a document;
26 storing information identifying a document in which a context occurs;
27 receiving a query containing an unspecified portion; and
28 identifying a plurality of matches for the query within the contexts.48.
29 A method of generating an index for satisfying an information need from a

- 1 48. A method of generating an index for satisfying an information need from a
 2 plurality of document stored on a computer-readable medium, comprising the
 3 steps of:
 4 receiving a document;
 5 identifying a context in the document;
 6 linguistically analyzing the context; selecting a term from the document;
 7 determining if there are more terms in the context to select, and if so,
 8 selecting another term from the context until there are no more
 9 terms to select;
 10 determining if there are more contexts to identify in the document,
 11 and if so, identifying another context in the document and repeating the
 12 term selecting step for each term in the context until there are no more
 13 contexts to identify;
 14 determining if there are more documents to receive, and if so,
 15 receiving the next document and repeating the context identifying and
 16 term selecting steps until there are no more documents to receive.
- 17 49. The method of claim 48, wherein linguistically analyzing the one or more
 18 contexts further comprises identifying category restrictions satisfied by the
 19 context.
- 20 50. The method of claim 48, further comprising the step of:
 21 storing information indicating the number of contexts stored for a term in
 22 a document, a document identifier associated with the document, and
 23 context identifiers indicating the location of the one or more contexts
 24 stored in the context array.
- 25
- 26 51. The method of claim 48, wherein the analyzed contexts are stored as finite state
 27 machines.
- 28 52. The method of claim 48, wherein the analyzed contexts are stored as graphs.
- 29 53. The method of claim 48, wherein the analyzed contexts are stored as trees.

- 1 54. The method of claim 48, wherein the category restrictions comprise
2 morphological features.
- 3 55. The method of claim 48, wherein the category restrictions comprise syntactic
4 features.
- 5 56. The method of claim 48, wherein the category restrictions comprise computer
6 programs.
- 7 57. An apparatus for fulfilling an information need employing an index stored on a
8 computer-readable medium and comprised of preanalyzed contexts of terms
9 appearing within a plurality of documents and information indicating category
10 restrictions that the terms and contexts satisfy, comprising:
11 memory means that stores computer-executable process steps; and
12 a processor that executes the process steps so as to (i) receive a query
13 comprised of one or more fully specified terms and an information need,
14 wherein the information need is represented by one or more at least
15 partially unspecified terms reflecting a category restriction; (ii) identify
16 preanalyzed contexts in the index that contain the one or more fully
17 specified terms and the one or more at least partially unspecified terms;
18 and (iii) locate one or matches for the information need within the
19 identified preanalyzed contexts.
- 20 58. An apparatus for fulfilling an information need employing an index stored on a
21 computer-readable medium and comprised of preanalyzed contexts of terms
22 appearing within a plurality of documents, information indicating category
23 restrictions that the terms and contexts satisfy, and identifiers of the documents
24 and contexts containing the terms, comprising:
25 memory means that stores computer-executable process steps; and
26 a processor that executes the process steps so as to (i) receive a query
27 comprised of one or more fully specified terms and an information need
28 and at least a partial restriction on the order that the one or more fully
29 specified terms and the information need may appear in a potential

1 matching preanalyzed context, wherein the information need is
2 represented by one or more at least partially unspecified terms reflecting a
3 category restriction; (ii) identify preanalyzed contexts in the index that
4 contain the one or more fully specified terms and the one or more at least
5 partially unspecified terms in the specified order; and (iii) locate one or
6 matches for the information need within the identified preanalyzed
7 contexts.

8 59. An apparatus for fulfilling an information need employing an index stored on a
9 computer-readable medium and comprised of preanalyzed contexts of terms
10 appearing within a plurality of documents, information indicating category
11 restrictions that the terms and contexts satisfy, and identifiers of the documents
12 and contexts containing the terms, comprising:

13 memory means that stores computer-executable process steps; and
14 a processor that executes the process steps so as to (i) receive a query
15 comprised of one or more fully specified terms and an information need
16 and at least a partial restriction on the order that the one or more fully
17 specified terms and the information need may appear in a potential
18 matching preanalyzed context, wherein the information need is
19 represented by one or more at least partially unspecified terms reflecting a
20 category restriction; (ii) identify preanalyzed contexts and documents in
21 the index that contain the one or more fully specified terms and the one or
22 more at least partially unspecified terms in the specified order by
23 converting the query into a Boolean expression and identifying context
24 identifiers satisfying the Boolean expression; and (iii) locate one or
25 matches for the information need within the identified preanalyzed
26 contexts by converting the query into a finite state machine and matching
27 the finite state machine against the identified preanalyzed contexts.

- 1 60. Computer-executable process steps stored on a computer-readable medium, the
2 computer-executable process steps to fulfill an information need, the computer-
3 executable process steps comprising:
4 code to receive a query containing an unspecified portion, the unspecified
5 portion including an unspecified term; and
6 code to identify one or more matches for the query within a body of
7 information stored on a computer-readable medium.
- 8 61. Computer-executable process steps stored on a computer-readable medium, the
9 computer-executable process steps to fulfill an information need, the computer-
10 executable process steps comprising:
11 code to store contexts for terms, wherein a context occurs in a document,
12 code to store information identifying a document in which a context
13 occurs;
14 code to receive a query containing an unspecified portion; and
15 code to identify one or more matches for the query within the contexts.
- 16 62. Computer-executable process steps stored on a computer-readable medium, the
17 computer-executable process steps to fulfill an information need, the computer-
18 executable process steps comprising:
19 code to identify a plurality of matches for a partially unspecified query;
20 and
21 code to rank a plurality of the matches or portions thereof.
- 22 63. Computer-executable process steps stored on a computer-readable medium, the
23 computer-executable process steps to fulfill an information need, the computer-
24 executable process steps comprising:
25 code to identify a plurality of results for a query, the results occurring
26 within documents; and
27 code to rank the plurality of results based on the content of a plurality of
28 documents in which a result is identified.